Application No. 09/893,301   Atty Docket No. INXT 1017-1

Amendments to the Specification

Under the heading **BRIEF DESCRIPTION OF THE DRAWINGS** on page 2 of the specification, please enter the following amendments:.

On page 2, after paragraph [0011], please add the following new paragraph:

[0011A]   Figure 7 depicts registering documents in a training set.

On page 7, please amend paragraph [0024] as follows:

[0024]   From the tables discussed above, the process of registering documents in a training set can be revisited by reference to Figure 7, this time for a training set of coded documents. A coded document 701 is received 702. A term vector (termCount in table 600) is created. The text of the document is stored in a table 601. When the training set has been loaded, term vectors of training set documents are compared to generate similarity scores 703. Many different measures of similarity can be practiced in accordance with the present invention; the present invention does not depend on the similarity measure used. From the similarity scores, k nearest neighbor similar document lists can be created for the documents 704, where k is a parameter set for the number nearest neighbors to process. Category scores (stored in table 606) are calculated, based on the nearest neighbors and editorial assignments 705. Next, precision and recall curves are constructed 705, using the nearest neighbor data. Category assignment thresholds are established 706 by analysis of the curves. This may be a manual or automatic process; the threshold setting process is not important to the present invention. Workfiles and data used to compile the curves and to set the thresholds are erased in the normal course of processing, or at least not reused. If documents are added to or deleted from the training set database, the nearest neighbors are reevaluated, similarity scores, curves and category scores recalculated, and adjustment of the category assignment thresholds is at least considered. Substantial effort is involved in updating the entire training set database.